ManageEngine
AD360

Adoptar Zero Trust para protegerse de las ciberamenazas de la IA generativa

Adoptar Zero Trust para protegerse de las ciberamenazas de la IA generativa

La IA generativa, tal y como la define el Foro Económico Mundial, se refiere a algoritmos que crean nuevos resultados a partir de los datos con los que se entrenan. El sistema de IA generativa no es como la IA tradicional, que sigue patrones y reglas predeterminados, sino que aprende de los datos sin instrucciones explícitas para crear nuevos contenidos como audio y arte. A medida que avanza la IA generativa, también plantea riesgos para la seguridad, ya que puede utilizarse para crear vídeos falsos convincentes o correos electrónicos de phishing, lo que podría dar lugar a desinformación, fraude y violaciones de la privacidad.

En este e-book, abordaremos la IA generativa y la seguridad informática, pero empezaremos por obtener más información sobre la IA generativa y cómo funciona.

El auge de la IA generativa

Echemos un vistazo a la historia de la IA generativa desde el pasado hasta el presente y más allá.

Las raíces de la IA generativa se remontan a los primeros días del machine learning. A finales de la década de 1950, científicos e investigadores empezaron a explorar la idea de utilizar algoritmos para crear nuevos datos. En aquella época, uno de los primeros ejemplos de IA generativa era la cadena de Markov, un modelo estadístico capaz de producir nuevas secuencias de datos a partir de la información que recibía.

Gracias a los avances en potencia de cálculo y algoritmos, la IA generativa ha progresado notablemente. Durante la década de 1980, los investigadores desarrollaron técnicas más sofisticadas, como las redes bayesianas y los modelos de Markov ocultos, para generar datos y modelar relaciones complejas.

En la década de 1990, el deep learning surgió como una nueva forma de machine learning que utiliza redes neuronales artificiales para obtener conocimientos a partir de los datos. Los modelos de deep learning comprendían relaciones entre puntos de datos mucho más complejas que los modelos anteriores, lo que dio lugar a una nueva ola de progreso de la IA.

La década de 2000 vio el nacimiento de las redes generativas adversarias (GAN), un revolucionario modelo de deep learning. Los GAN son redes neuronales que compiten entre sí: una red genera nuevos datos, mientras que la otra distingue entre los datos reales y los generados. Este proceso ayuda a la red generadora a aprender a crear datos realistas y a distinguir entre datos reales y falsos.

En los últimos años se han producido grandes avances en la IA generativa, como el modelo GPT-3 de OpenAI. Estos modelos se utilizan para crear arte y música, desarrollar nuevos productos y mejorar la asistencia médica. A medida que avanza la tecnología y aumenta el acceso a los datos, la IA generativa se expande y evoluciona, ofreciendo nuevas oportunidades de innovación y descubrimiento.

La IA generativa ha acaparado interés y titulares recientemente, y encierra un inmenso potencial de crecimiento exponencial en el futuro. Los contenidos generados por IA serán aún más sofisticados y creativos a medida que continúe la investigación, difuminando los límites entre los contenidos creados por humanos y los creados por IA.

¿Cómo funciona la IA generativa?

La IA generativa utiliza el machine learning y las redes neuronales para encontrar patrones en los datos. Los modelos de IA aprenden a partir de grandes cantidades de datos que reciben durante el entrenamiento. Puede ser texto, código, gráficos o cualquier cosa relevante para la tarea.

A partir de los datos de entrenamiento, un modelo de IA analiza los patrones y las relaciones dentro de los datos para comprender las reglas subyacentes que controlan el contenido. A medida que aprende, el modelo de IA ajusta sus parámetros, lo que le permite simular mejor los contenidos generados por humanos. Cuantos más contenidos generen los modelos de IA, más sofisticados y convincentes serán los resultados.

En pocas palabras, los usuarios suelen interactuar con la IA generativa proporcionando algún tipo de indicación, que puede estar en cualquier formato que el sistema pueda procesar. En respuesta a la indicación, se devuelven nuevos contenidos al usuario.

Ejemplos de IA generativa

Estas son algunas de las interfaces de IA más populares.

ChatGPT: Está entrenado para conversar en lenguaje natural. Puede escribir historias, ensayos, poemas y recetas, así como entablar conversaciones de ida y vuelta con usted.

DALLE-E: También utiliza el procesamiento del lenguaje natural para generar nuevas imágenes. DALL-E incorpora texto descriptivo para generar imágenes fotorrealistas basadas en el mensaje. También se pueden generar imágenes con distintas perspectivas y estilos.

BARD: Al igual que ChatGPT, Bard funciona con la tecnología de IA de Google. El sistema de IA se entrena con una gran recopilación de documentos de texto y código, y es capaz de generar texto, traducir idiomas, redactar contenidos creativos y responder preguntas.

Midjourney: Crea imágenes realistas y artísticas basadas en indicaciones de texto utilizando un bot de Discord. Usted puede editar, ampliar y descargar sus creaciones. También ayuda a crear descripciones para las imágenes basadas en las entradas de los usuarios.

Un arma de doble filo

Cualquier industria u organización puede beneficiarse de la IA generativa para aumentar la productividad, automatizar tareas, permitir nuevas formas de creación y desarrollar datos sintéticos para entrenar modelos de IA. Aunque esta tecnología tiene sus ventajas, hay que tener en cuenta algunos aspectos. Analicemos los aspectos positivos y negativos de la IA generativa.

El poder de la IA generativa

- Automatiza y acelera la realización de tareas para aumentar la productividad.
- Elimina o reduce las barreras de habilidad y tiempo para generar contenidos y desarrollar aplicaciones creativas.
- Permite explorar y analizar datos complejos.
- Ayuda a crear datos sintéticos para entrenar y mejorar otros sistemas de IA.

No olvide los riesgos de la IA generativa

- La IA generativa puede crear deepfakes convincentes, que pueden utilizarse para difundir información errónea, difamar a personas o dañar reputaciones.
- El uso de datos confidenciales o privados para entrenar modelos generativos de IA puede dar lugar a fugas involuntarias de datos a través del contenido generado.
- Los hackers pueden aprovechar las vulnerabilidades de las aplicaciones de IA generativa para obtener acceso no autorizado, manipular datos o perturbar el sistema.
- Unas pruebas inadecuadas, unos controles de acceso deficientes y una manipulación incorrecta de los datos pueden provocar violaciones de seguridad en los modelos generativos de IA.

En el informe Top Cybersecurity Threats In 2023 de Forrester, las aplicaciones de IA como ChatGPT figuran entre las amenazas emergentes, y la analista de Gartner Avivah Litan destaca cinco riesgos principales asociados a la IA generativa: información fabricada, deepfakes, privacidad de datos, problemas de derechos de autor y seguridad informática.

En términos de seguridad informática, la IA generativa plantea una preocupación central porque podría proporcionar a los atacantes grandes poderes para desarrollar exploits maliciosos y llevar a cabo ataques cibernéticos más efectivos. Entonces, ¿cómo utilizan los hackers la IA generativa en sus ataques?

Cómo se utiliza la IA generativa en los ciberataques

A medida que la IA generativa se generaliza, también aumenta el potencial de los ataques cibernéticos que aprovechan esta tecnología. El uso de la IA generativa por parte de los delincuentes informáticos podría plantear graves amenazas para la seguridad y la integridad de los datos de las organizaciones. Exploremos algunos de los riesgos de la IA generativa en la seguridad informática:

Phishing de credenciales: la IA se utiliza en el phishing de credenciales para crear correos electrónicos que parecen proceder de una empresa legítima o sitios web falsos que parecen reales. De este modo, es más probable que los usuarios revelen información confidencial o introduzcan sus credenciales.

Explotación de endpoints: Esto implica el uso de IA para automatizar el proceso de encontrar y explotar vulnerabilidades, crear vectores de ataque sofisticados y evadir la detección.

Correo electrónico corporativo comprometido (BEC): Al utilizar IA en los ataques BEC, los correos electrónicos parecen más reales y convincentes. La IA podría, por ejemplo, analizar los patrones de correo electrónico de un director general y crear un mensaje que imite su estilo de escritura. Esto aumenta las posibilidades de que el destinatario confíe en el correo electrónico y siga las instrucciones.

Creación de malware: La IA se utiliza en la creación de malware para crear código más difícil de detectar por los antivirus. Por ejemplo, la IA puede utilizarse para automatizar el desarrollo de programas maliciosos. Analizando muestras de malware existentes y aprendiendo de su comportamiento, los algoritmos de IA pueden crear malware complejo y polimórfico que evoluciona y se adapta constantemente para escapar a la detección del software antivirus.

Para contrarrestar los ataques cibernéticos de IA generativa, las organizaciones deben adoptar el modelo de seguridad de Zero Trust. Esta estrategia hace hincapié en el principio de "nunca confíes, verifica siempre", y exige autenticación y autorización constantes para cada usuario, dispositivo o aplicación que intente acceder a la red o a los datos.

Adopción de Zero Trust

En la publicación especial 800-207 del NIST, Zero Trust se describe como un paradigma de ciberseguridad que desplaza el centro de atención de la protección de grandes redes a la protección de individuos o pequeños grupos. Según su ubicación física o en la red en Zero Trust, no se concede confianza implícita a los activos ni a las cuentas de usuario. En cambio, la autenticación y la autorización implican pasos finitos antes de poder acceder a un recurso de la empresa.

Con Zero Trust, las organizaciones pueden reducir al mínimo sus superficies de ataque, minimizar el impacto de las violaciones de seguridad y mejorar su postura general de seguridad mediante la verificación continua de las identidades de usuarios y dispositivos.

En 2023, una encuesta realizada a responsables de seguridad de 31 países, que representaban a casi todas las industrias y al sector público, reveló que solo el 28% de las organizaciones implementaban una solución completa de Zero Trust para remediar los riesgos derivados de los ataques cibernéticos. Esto significa que las organizaciones deben dar prioridad a la implementación de estrategias integrales de confianza cero para reforzar su postura de seguridad y proteger los datos sensibles de las amenazas informáticas sofisticadas.

El enfoque Zero Trust implica las siguientes medidas de seguridad:

Gestión de accesos e identidades (IAM): La IAM es crucial en la seguridad de la información, ya que determina quién puede acceder a qué recursos y en qué condiciones. Cada solicitud se verifica y autoriza en función de múltiples factores, como la identidad, el contexto, el dispositivo, la ubicación y el comportamiento. En la IAM, la autenticación, la autorización y la gestión del acceso son los componentes clave.

Monitoreo y auditoría continuas: Para detectar actividades sospechosas o anomalías, confianza cero monitorea el comportamiento de usuarios y dispositivos. Implica la recopilación y el análisis de datos procedentes de diversas fuentes, como el tráfico de red, los logs de actividad de los usuarios y las soluciones de seguridad para endpoints. Proporciona a las organizaciones datos en tiempo real y pone en práctica procedimientos de seguridad como la respuesta a incidentes, la evaluación de amenazas, el análisis forense de equipos y bases de datos y el análisis de la causa raíz.

Análisis del comportamiento de entidades y usuarios (UEBA): Mediante la implementación de soluciones UEBA, se detectan anomalías como intentos de acceso no autorizados y transferencias de datos inusuales. Con las herramientas UEBA que priorizan la intensidad de la amenaza, las organizaciones pueden decidir centrarse primero en los problemas de alto riesgo. Además, UEBA ayuda a detectar una gama más amplia de ataques, como ataques DDoS, ataques de fuerza bruta, exfiltración de datos y ataques internos.

Respuesta a incidentes (IR): La IR implica identificar los ataques, comprender su gravedad, priorizarlos, investigarlos y mitigarlos, restablecer las operaciones y evitar que se repitan. Las organizaciones pueden minimizar el impacto de los incidentes, reducir el tiempo de inactividad y proteger su reputación aplicando planes de IR.

Inteligencia sobre amenazas: El objetivo de la inteligencia sobre amenazas es recopilar, analizar y compartir información sobre las amenazas a la seguridad de los servicios, redes y datos de una organización. Esto incluye qué están haciendo los hackers, cómo lo están haciendo y qué vulnerabilidades están explotando. Gracias a la inteligencia sobre amenazas, los equipos de seguridad pueden comprender mejor las amenazas y responder de forma más proactiva para proteger los activos de su organización.

Cómo ManageEngine puede ayudar a aplicar un modelo de Zero Trust

Las redes de Zero Trust son una capa adicional de seguridad contra los ataques cibernéticos de IA generativa. Por ejemplo, en el phishing de credenciales, Zero Trust sirve como barrera que impide a los delincuentes informáticos avanzar en sus objetivos maliciosos. Cuando se requiere MFA para acceder a una cuenta, a un delincuente informático le resultará difícil entrar en ella. Al adherirse al paradigma de confianza cero, incluso si una cuenta se ve comprometida y vinculada a múltiples aplicaciones, la capacidad de un atacante para acceder a otras aplicaciones permanece restringida. Además, el monitoreo periódico del comportamiento de los usuarios minimiza las amenazas externas e internas al detectar a tiempo posibles amenazas.

Para aplicar un modelo de Zero Trust y reforzar su postura de seguridad informática, ManageEngine proporciona estas funciones:

Autenticación multifactor (MFA): Añadir una segunda capa de autenticación a las contraseñas existentes ayuda a frustrar el acceso no autorizado, incluso si los atacantes consiguen obtener ciertas credenciales mediante ataques de phishing. Con ManageEngine AD360, una solución IAM empresarial, usted puede implementar MFA para:

- Equipos Windows, macOS y Linux
- Los mejores proveedores de VPN como Fortinet, Cisco AnyConnect, Pulse y muchos más.
- Endpoints compatibles con autenticación RADIUS como Citrix Gateway, VMWare Horizon y Microsoft Remote Desktop Gateway (RDP).
- Inicios de sesión en Outlook Web Access (OWA)

Hay 19 factores de autenticación diferentes, entre los que se incluyen la huella dactilar, el identificador facial, la verificación por correo electrónico, Duo Security, Microsoft Authenticator, Google Authenticator y YubiKey Authenticator. Usted puede consultar la lista completa de autenticadores compatibles aquí.

Autenticación sin contraseña: La autenticación sin contraseña elimina la necesidad de crear contraseñas, lo que reduce el riesgo de robo de credenciales. ManageEngine AD360 ofrece un SSO empresarial que proporciona a los usuarios un acceso fluido con un solo clic:

- Aplicaciones compatibles con SAML
- Aplicaciones compatibles con OAuth y OpenID Connect
- Aplicaciones personalizadas

Usted también puede controlar quién accede a qué aplicaciones en la nube creando políticas basadas en dominios, OU y grupos.

Principio del menor privilegio: Conceder a cada usuario acceso solo a los recursos esenciales y nada más, delimitará el riesgo si las credenciales del usuario se ven comprometidas en un incidente de seguridad. ManageEngine AD360 proporciona:

- Informes predefinidos, que ayudan a ver y gestionar eficientemente tanto los permisos de acceso nuevos como los heredados a los archivos compartidos.
- Plantillas que ayudan a los administradores a conceder solo los permisos necesarios a los nuevos empleados. Estas plantillas rellenan automáticamente los permisos necesarios, como pertenencia a grupos, permisos de servidor de archivos, etc., en función de su rol o designación.
- Gestión automatizada de permisos con límite de tiempo que permite asignar temporalmente usuarios a grupos específicos y conceder permisos al servidor de archivos.

Análisis del comportamiento de usuarios y entidades: En las soluciones UEBA, el comportamiento anómalo de los usuarios se detecta analizando patrones y desviaciones del comportamiento normal. Además, se pueden configurar alertas en tiempo real para notificar a los equipos de seguridad actividades sospechosas, lo que les permite investigar y responder a los incidentes con rapidez. Con ManageEngine Log360, una solución SIEM unificada, usted puede:

- Resolver rápidamente las amenazas de alto riesgo asignando puntuaciones de riesgo a los distintos eventos de seguridad.
- Detectar anomalías con mayor precisión y reducir los falsos positivos agrupando a los usuarios en función de sus comportamientos y estableciendo alertas para las desviaciones de las líneas de base establecidas.
- Reducir el tiempo de espera recibiendo alertas de seguridad en tiempo real basadas en el comportamiento a través de SMS o correo electrónico.

Detección y respuesta a incidentes: Para minimizar los efectos de un incidente de seguridad, una organización debe contar con un proceso efectivo de gestión de incidentes. El sistema de gestión de incidentes ManageEngine Log360 ayuda a los equipos de seguridad a responder a las amenazas de seguridad informática de forma rápida y eficiente. Con esta solución, usted puede:

- Mejorar la gestión de incidentes con un dashboard de incidentes que muestra los incidentes activos y sin resolver junto con los incidentes recientes y críticos, realizar un control de las métricas clave y priorizar la resolución de incidentes para un rendimiento óptimo del SOC.

- Identificar patrones de ataque como inyección SQL, denegación de servicio y ataques de firewall mediante más de 30 reglas de correlación predefinidas.
- Priorizar a los incidentes de alto riesgo y agilizar la resolución de amenazas con un sistema de alertas en tiempo real que incluye más de 1.000 criterios de alerta predefinidos.
- Automatizar su respuesta a las amenazas a la seguridad con una gestión de flujos de trabajo que aborde al instante las amenazas internas, las cuentas comprometidas y los intentos de exfiltración de datos.
- Integrar con software de mesa de ayuda externo como ManageEngine ServiceDesk Plus, Zendesk, ServiceNow y Kayako para generar tickets cuando se activen las alertas.

Inteligencia sobre amenazas: En la era de los ataques cibernéticos de IA generativa, las organizaciones necesitan incorporar inteligencia de amenazas en su estrategia de seguridad para mantenerse a la vanguardia. ManageEngine Log360 le ayuda a:

- Obtener información sobre amenazas de fuentes basadas en STIX/TAXII como Hail A TAXII, AlienVault OTX o fuentes personalizadas.
- Activar automáticamente flujos de trabajo para bloquear permanentemente las IP incluidas en listas de bloqueo añadiéndolas al firewall.
- Proteger los datos detectando las comunicaciones salientes, notificando a los analistas las IP, dominios o URL maliciosos y bloqueándolos rápidamente.
- Mitigar los falsos positivos mediante un sistema de respuesta a eventos en tiempo real que diferencia las actividades sospechosas de las legítimas.

Proteger el futuro

A medida que avanza la IA, también evolucionan los riesgos dirigidos a estas tecnologías. La amplia disponibilidad de herramientas como ChatGPT y Google Bard ha otorgado a los atacantes la capacidad de elevar rápidamente la complejidad de los ataques aparentemente de la noche a la mañana. Los profesionales de TI deben adoptar estrategias proactivas para proteger a la organización de estas amenazas.

Para combatir las aplicaciones maliciosas de la IA generativa, es crucial desarrollar e implementar continuamente defensas sólidas, mejorar las funciones de detección y mantenerse alerta ante las amenazas emergentes, antes de convertirse en la próxima víctima de un ataque generado por IA.

Productos relacionados

Log360 | ADManager Plus | ADAudit Plus | ADSelfService Plus

Exchange Reporter Plus | RecoveryManager Plus

ManageEngine
AD360

ManageEngine AD360 es una solución unificada de gestión de accesos e identidades que ayuda a gestionar las identidades, proteger el acceso y garantizar el cumplimiento. Viene con potentes funciones como la gestión automatizada del ciclo de vida de las identidades, SSO seguro, MFA adaptable, flujos de trabajo basados en la aprobación, protección contra las amenazas a la identidad impulsada por UBA e informes de auditoría históricos para AD, Exchange Server y Microsoft 365. La interfaz intuitiva y las potentes funciones de AD360 la convierten en la solución ideal para todas sus necesidades de IAM, incluido el fomento de un entorno de Zero Trust.

$ Cotización ⬇ Descargar